# RACE TO THE SPACE

*SUFIA FATHIMA*

*30 SEPTEMBER, 2022*

IBM Developer

SKILLS NETWORK

# OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
  - Visualization – Charts
  - Dashboard
- Discussion
  - Findings & Implications
- Conclusion

# EXECUTIVE SUMMARY

- Data collection using API
- Data collection with web scraping
- Data wrangling
- Exploratory Data Analysis with SQL
- Exploratory Data Analysis with Data Visualization
- Interactive visual analysis with Folium
- Machine Learning Prediction

# INTRODUCTION

- This is a report on analysis performed on the rocket launch by company SpaceY.
- It consists of methodology involving data collection, data wrangling, exploratory data analysis, data visualization, model development, model evaluation, and results.
- It consists of prediction of the first stage of the SpaceX Falcon 9 rocket’s landing and in turn determine the cost of the launch.
- Hence, with the help of data science we can train the model to depict if the first stage can be reused.

# METHODOLOGY

- Data collection methodology:

  Get requests to the Space X API and web scraping from Wikipedia

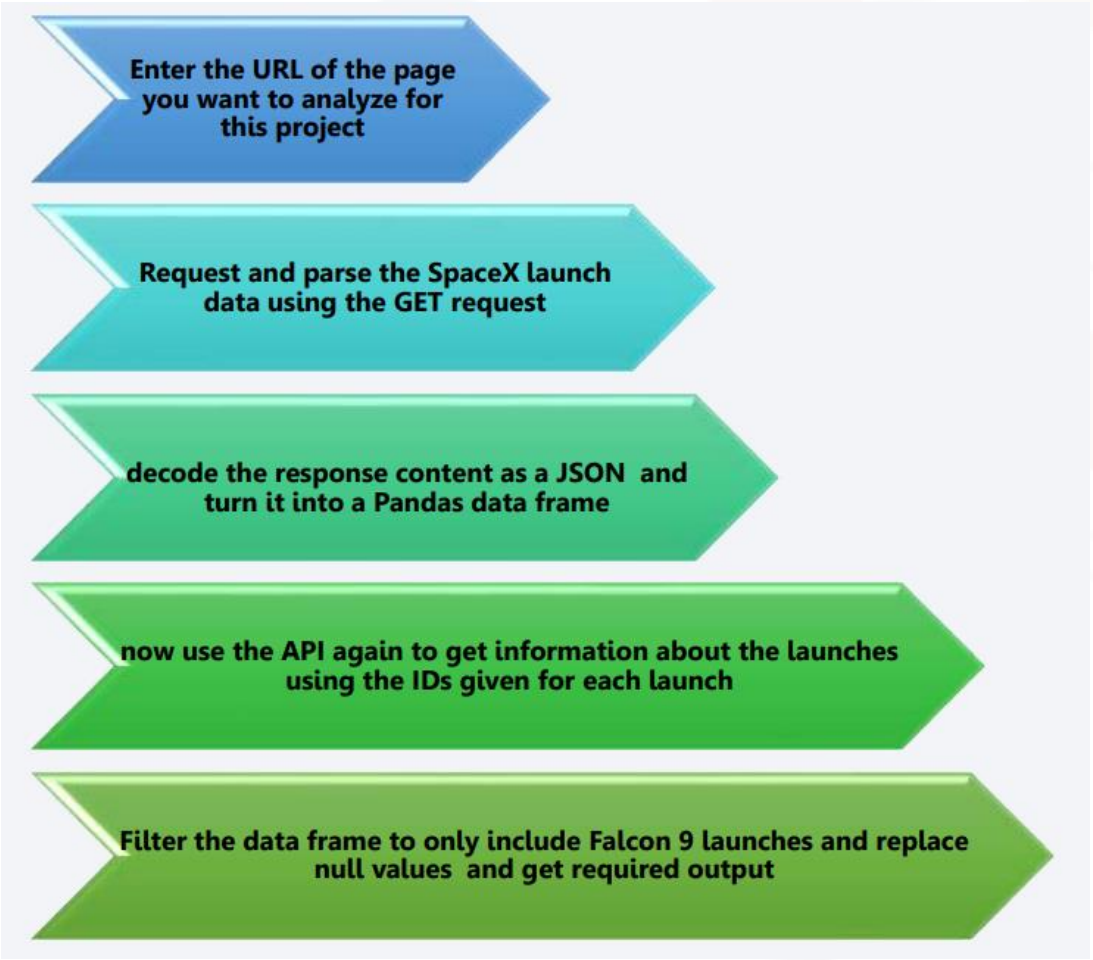
- Perform data wrangling:

  Clean the data

- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

  Creating best Machine Learning model.

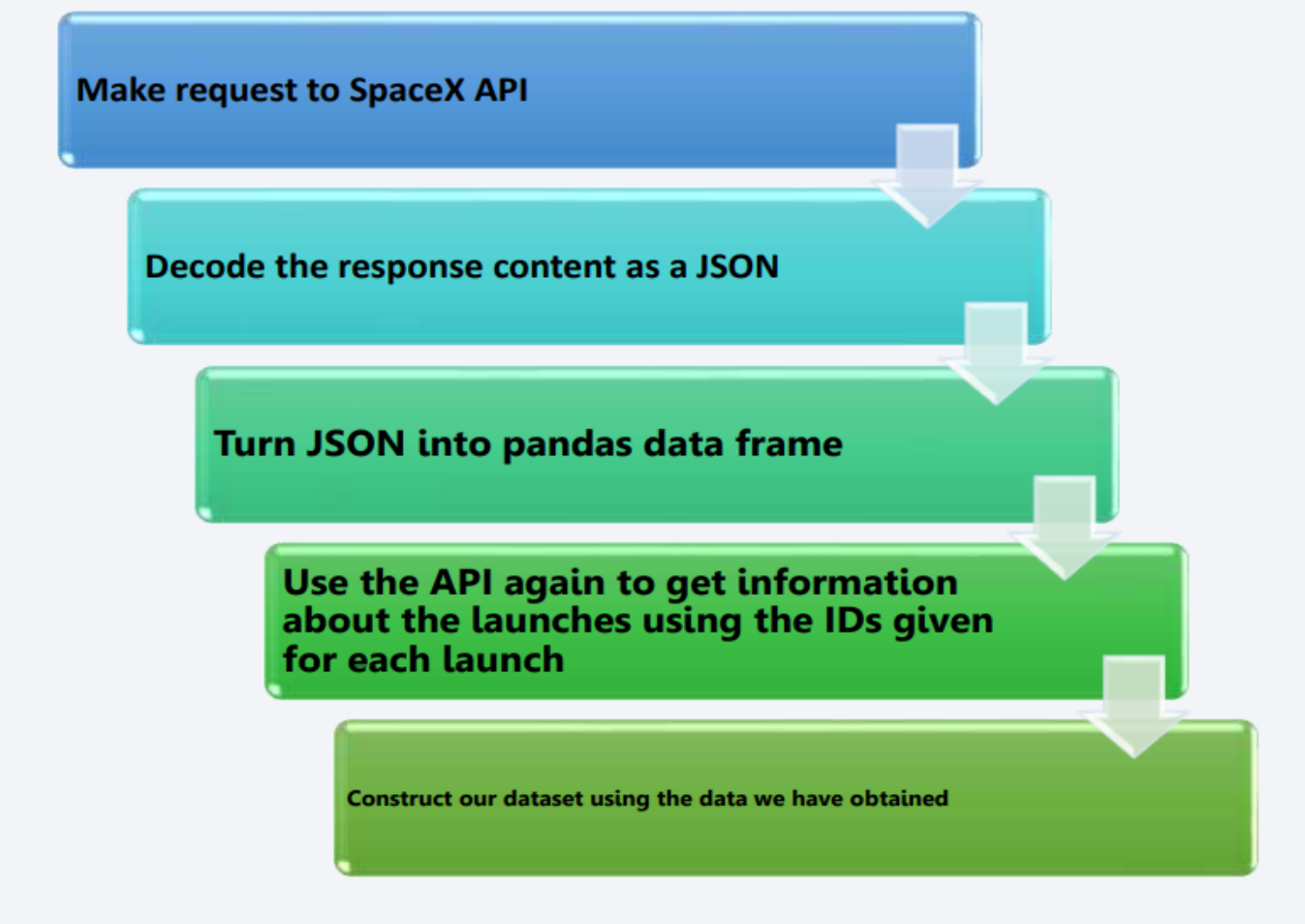
# DATA COLLECTION




The data sets are collected by:

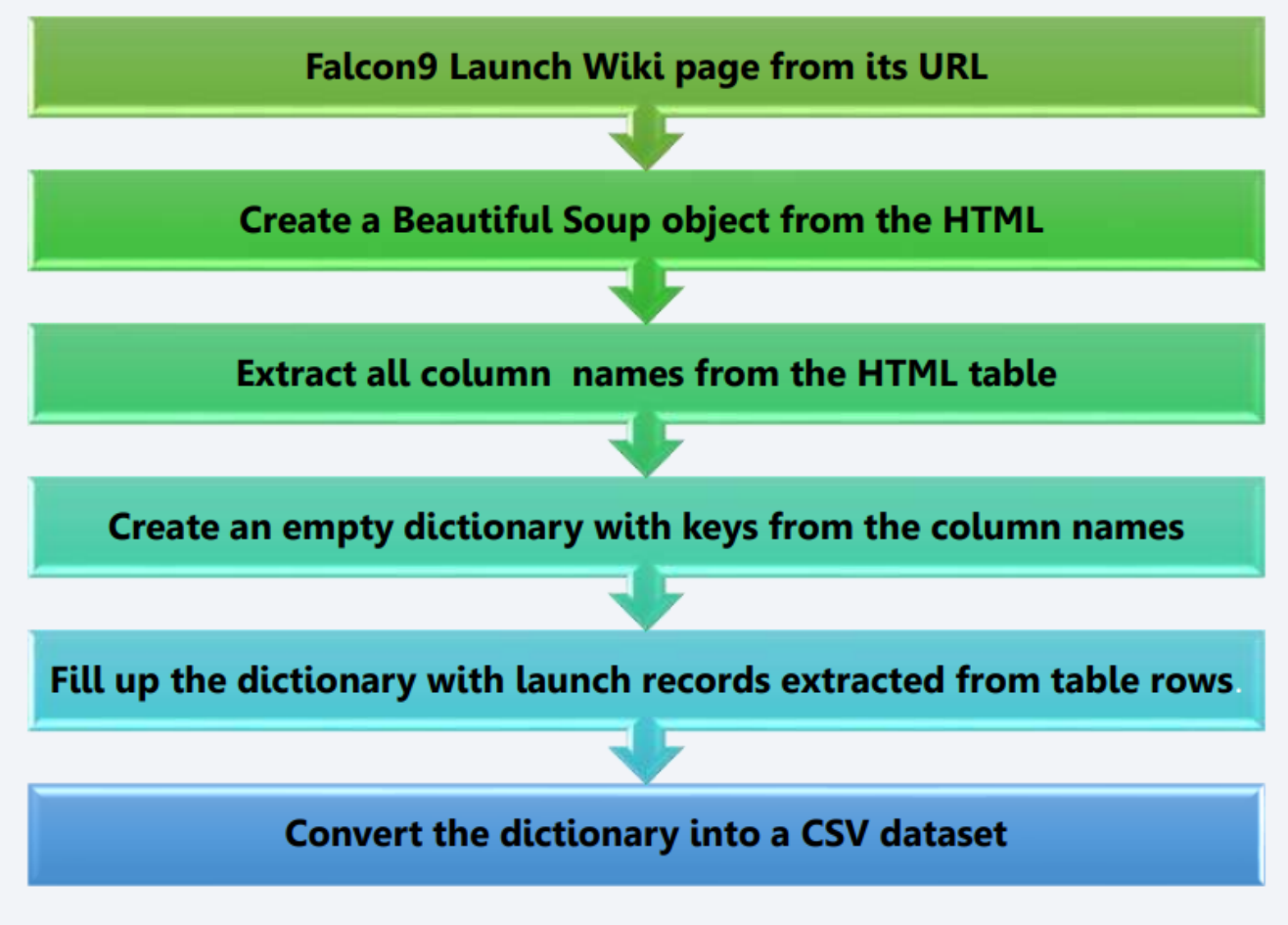
- SpaceX API request.
- Web Scraping

# DATA COLLECTION – SpaceX API




GITHUB LINK:
[Capstone-DataScience/Notebook1.ipynb at master · Supercalifragilisticexpialidocious19/Capstone-DataScience (github.com)]

# DATA COLLECTION – Web Scraping



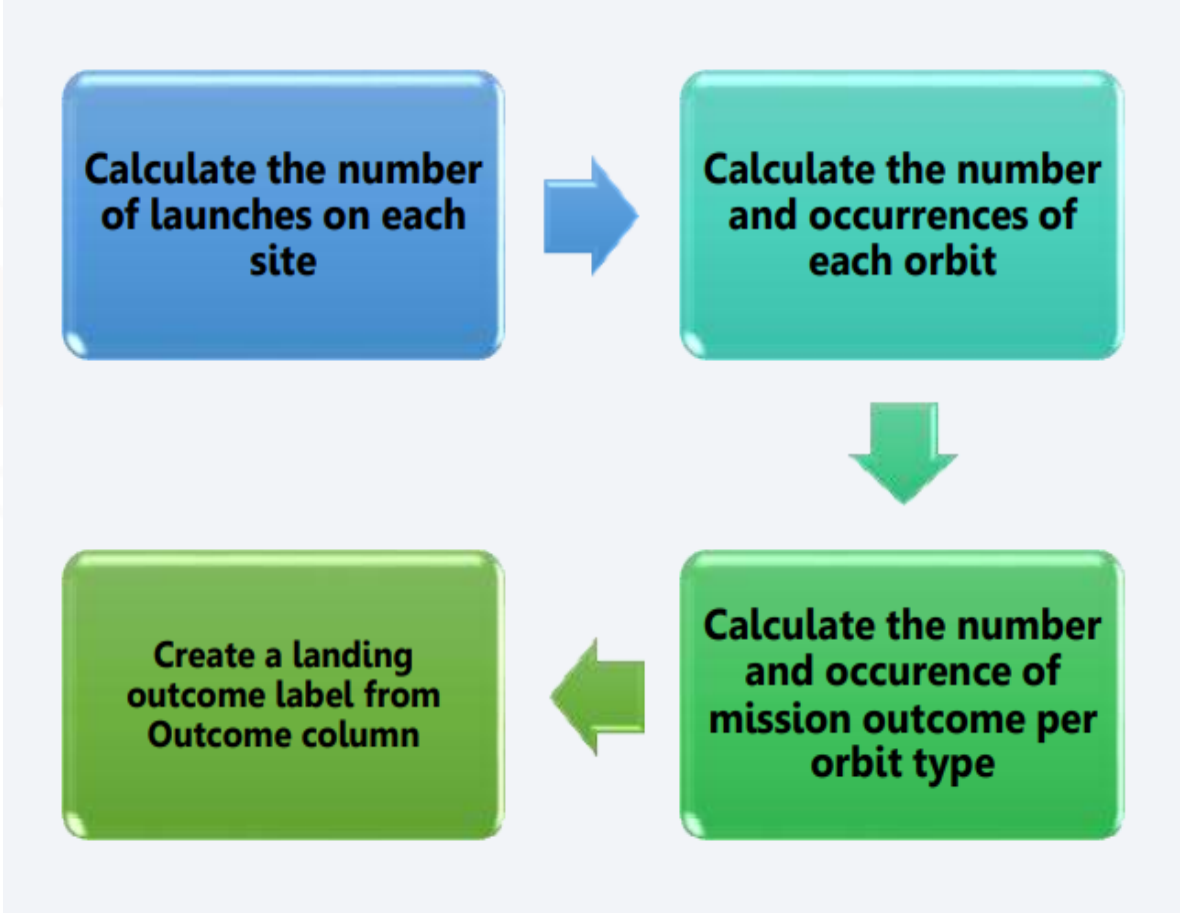

GITHUB LINK:

[Capstone-DataScience/data collection with web scraping.ipynb at master · Supercalifragilisticexpialidocious19/Capstone-DataScience (github.com)]

# DATA WRANGLING

- Data wrangling is the process in which data is transformed from its raw form to a more usable and understandable format.
- The flow chart is given as follows:
- The GITHUB link is as follows:

[Capstone-DataScience/Data Wrangling.ipynb at master · Supercalifragilisticexpialidocious19/Capstone-DataScience (github.com)]



# EDA with Data Visualization

Types of Charts Used :

- Scatter plot - Flight Number vs Payload Mass, Flight Number vs Launch Sites , Payload and Launch Sites , Flight Number and Orbit Type, Payload and Orbit Type
- Bar chart – Success rate of each orbit
- Line plot – success rate and Date

EDA with Data Visualization complete notebook link is given below:

Capstone-DataScience/EDA with Data Visualisation.ipynb at master · Supercalifragilisticexpialidocious19/Capstone-DataScience (github.com)

# EDA with SQL

- **Summary of SQL queries that were used:**

- Display the names of the unique launch sites in the space mission.
- Display 5 records where launch sites begin with the string ‘CCA’.
- Display the total payload mass carried by boosters launched by NASA (CRS).
- Display average payload mass carried by booster version F9 v1.1 .
- List the date when the first successful landing outcome in ground pad was achieved.
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000.

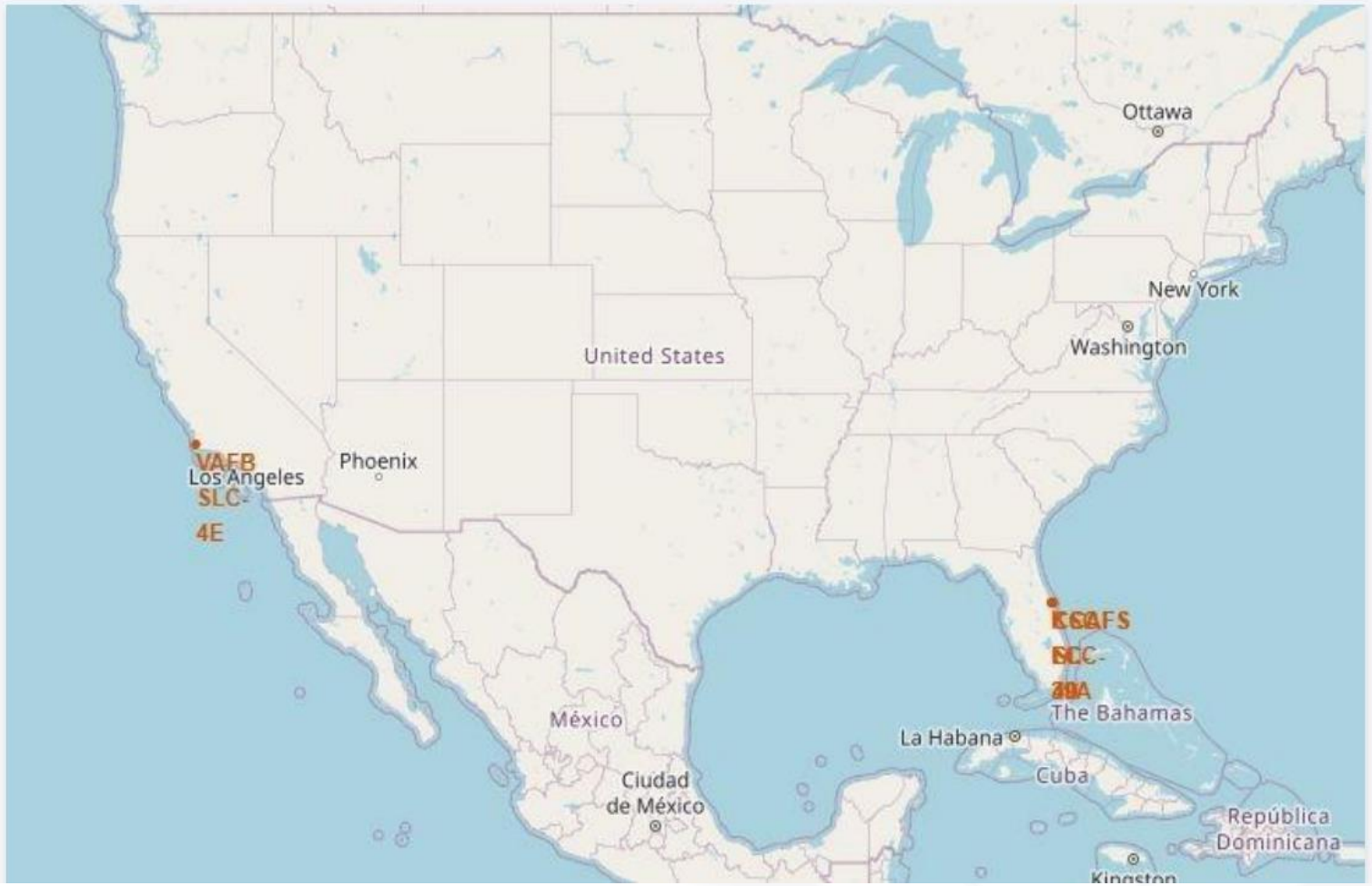
- List the total number of successful and failure mission outcomes.
- List the names of the booster versions which have carried the maximum payload mass. Use a subquery .
- List the failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015.
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order.

# Build an Interactive Map with Folium

- Folium Markers were used to show the Space X launch sites and their nearest important landmarks like railways, highways, cities and coastlines.
- Polylines were used to connect the launch sites to their nearest land marks.
- Red represents rocket launch failures
- Green represents the successes.
- The GITHUB link is as follows:

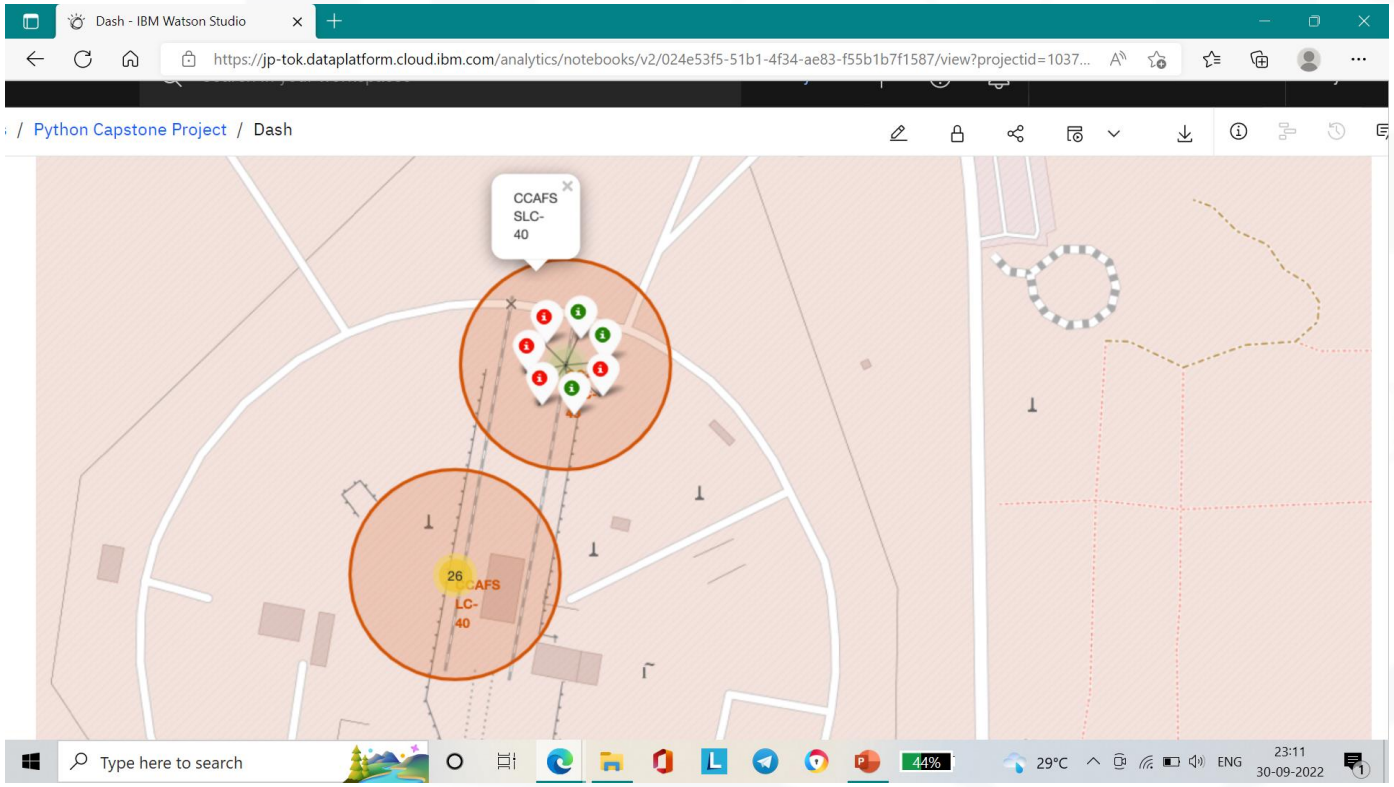
Capstone-DataScience/Dash.ipynb at master · Supercalifragilisticexpialidocious19/Capstone-DataScience (github.com)

# Launch site Locations




- The left map shows all *SpaceX* launch sites, and the right map also shows that all launch sites are in the United States.
- As can be seen on the map, all launch sites are near the coast.

# Success/failed launches for each site on the map - Insights




- We can easily identify which launch sites have relatively high success rates, by observing the color-labeled markers in marker clusters.

# Launch site distance to Landmarks






- Are launch sites in close proximity to railways? No
- Are launch sites in close proximity to highways? No
- Are launch sites in close proximity to coastline? Yes
- Do launch sites keep certain distance away from cities? Yes

# Dashboard with Plotly Dash

- Pie charts and scatter charts were used to visualize the launch records of Space X.
- These charts displayed the rocket launch success rate per launch site. We are were able to get an understanding of the factors that may have been influencing the success rate at each site. Such as the payload mass and booster versions.
- Successful launches were represented by 1 while failures were represented by 0.

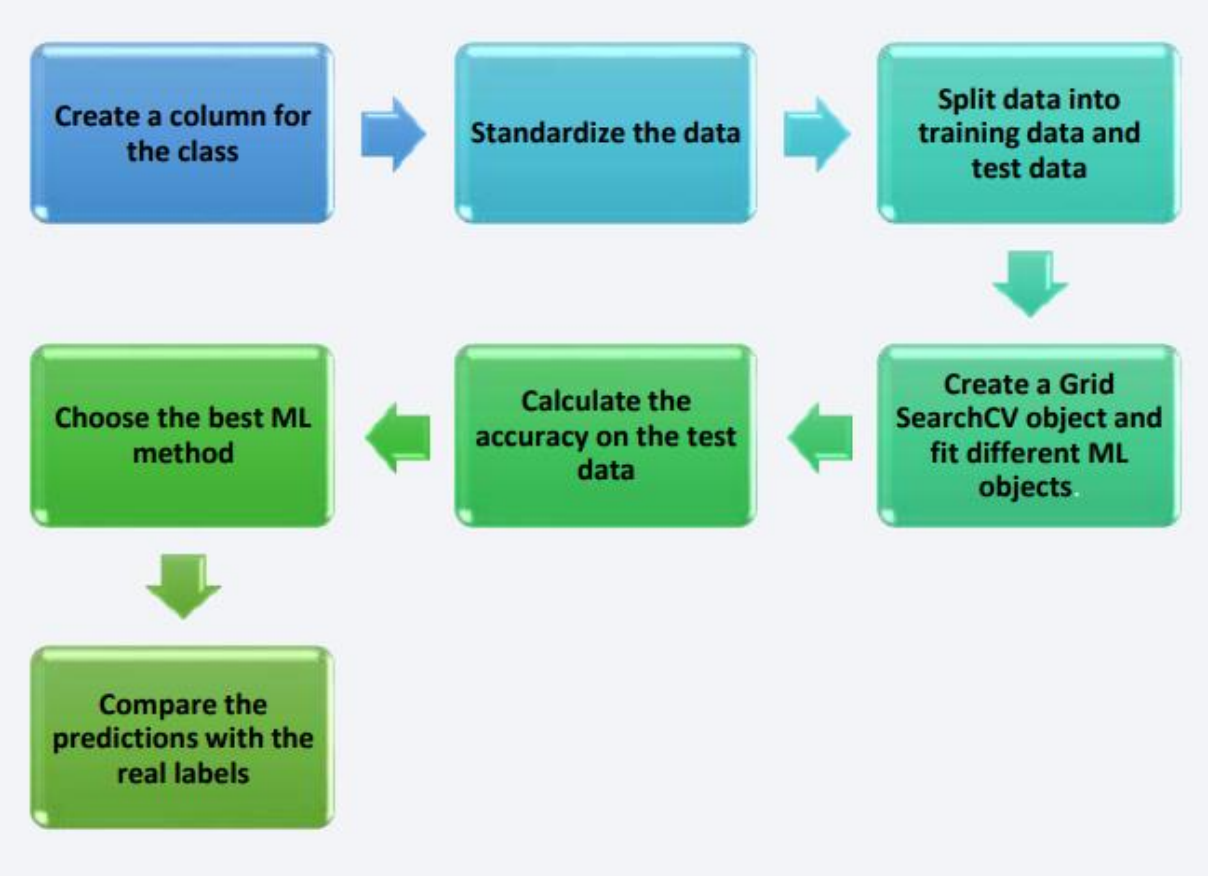
# Predictive Analysis (Classification)

- Scikit-learn is Machine Learning library that was used for predictive analysis.
- The following took place:

Created a machine learning pipeline to predict if the first stage will land given the data.

- The GITHUB link is as follows:

[Capstone-DataScience/Machine Learning Prediction.ipynb at master · Supercalifragilisticexpialidocious19/Capstone-DataScience (github.com)]



# RESULTS

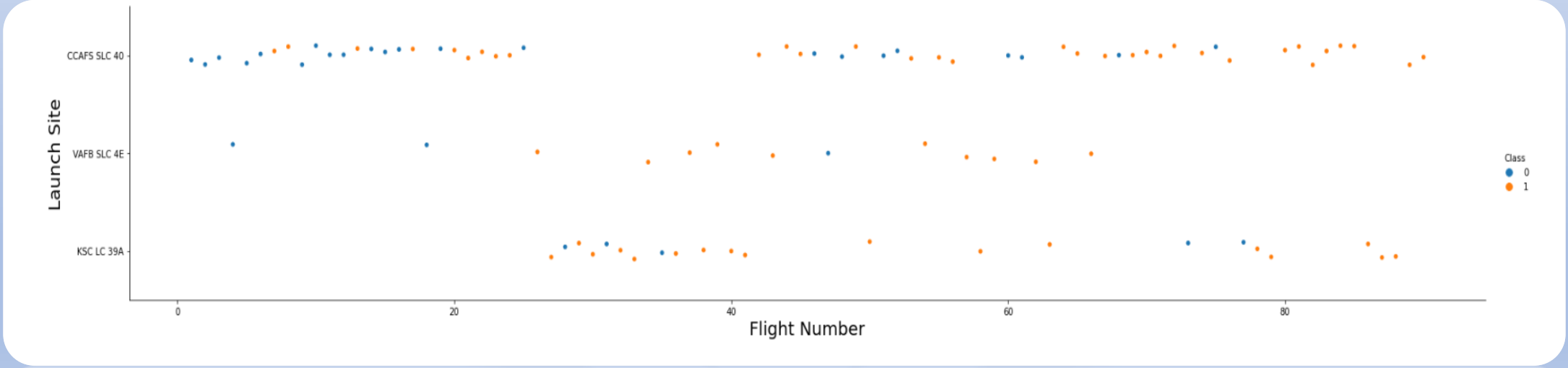
The exploratory data analysis has shown us that successful landing outcomes are somewhat correlated with flight number. It was also apparent that successful landing outcomes have had a significant increase since the year 2015.

- All launch sites are located near the coast line. Perhaps, this makes it easier to test rocket landings in the water.
- Sites are also located near highways and railways. This may facilitate transportation of equipment and research material.
- The machine learning were able to predict the landing success of rockets with an accuracy score of 83.33%.

# INSIGHTS DRAWN FROM EDA

## 1. Flight Number vs. Launch Site

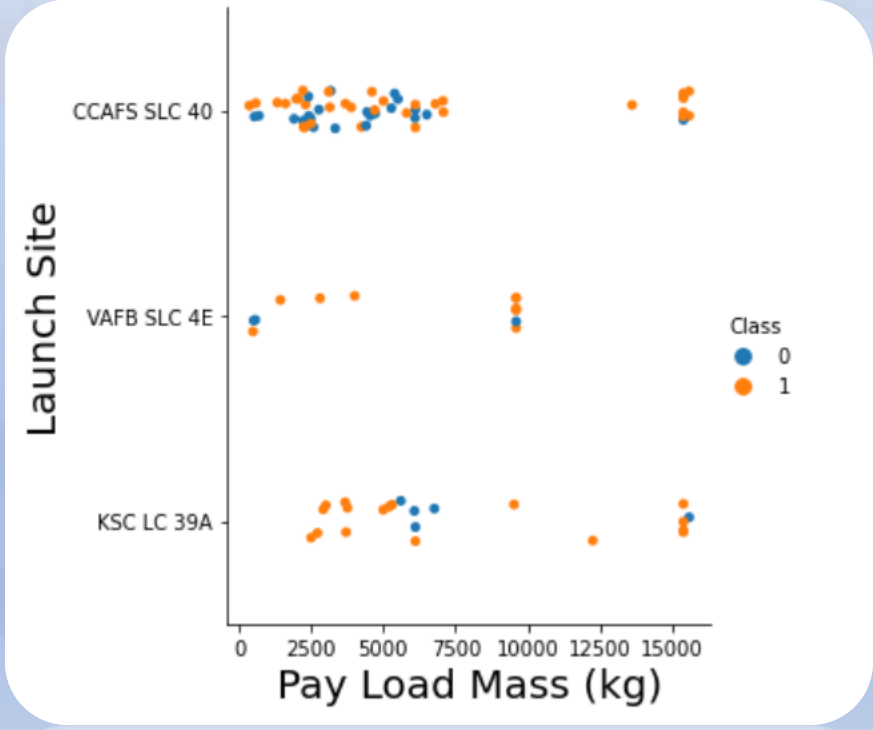
It appears that there were more successful landings as the flight numbers increased. launch site CCAFS SLC 40 had the most number of landing.



# INSIGHTS DRAWN FROM EDA

## 2. Payload VS Launch Site

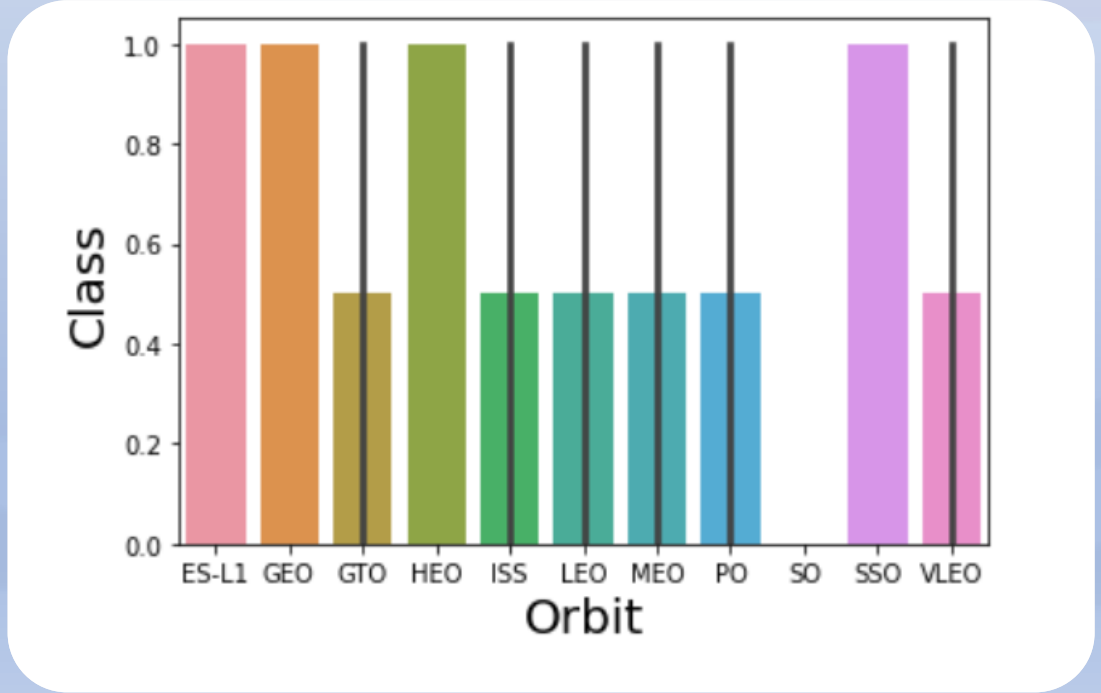
Now if you observe Payload Vs. Launch Site scatter point chart you will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).



# INSIGHTS DRAWN FROM EDA

## 3. Success Rate vs. Orbit Type

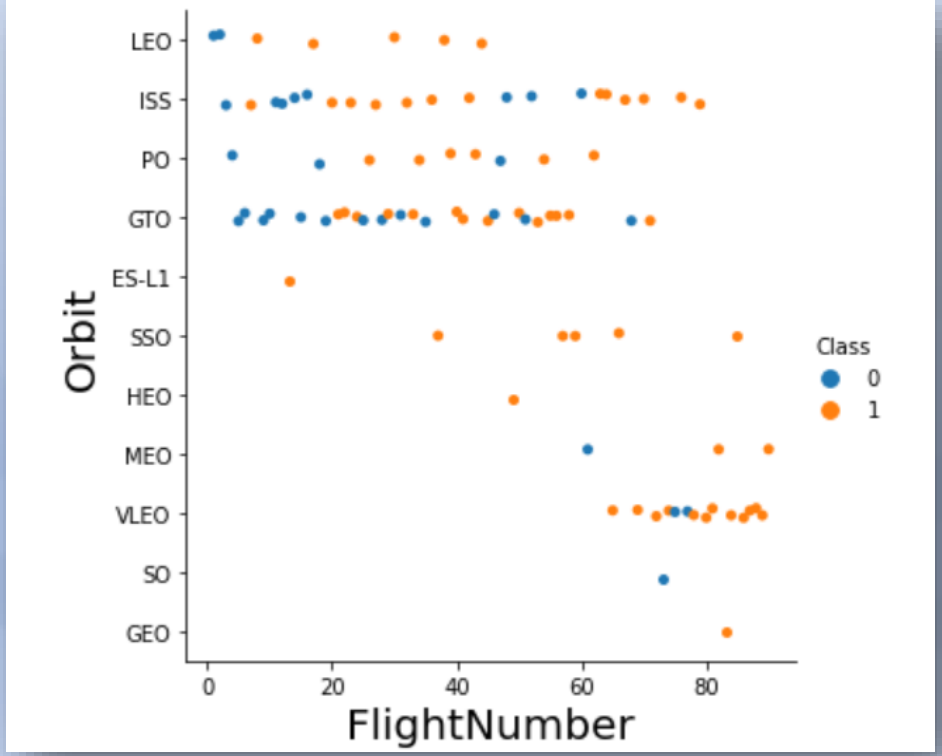
The highest success rate orbits are:  ES-L1, GEO, SSO and HEO.



# INSIGHTS DRAWN FROM EDA

## 4. Flight Number vs. Orbit Type

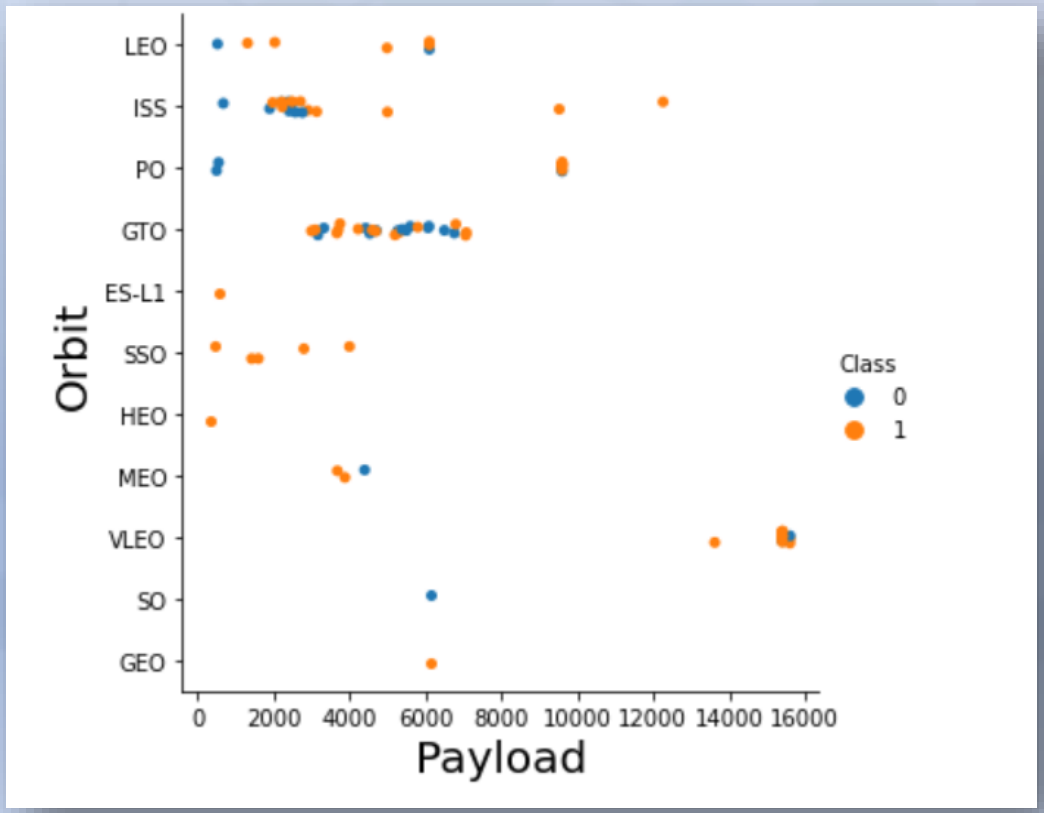
It is observed that in the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.



# INSIGHTS DRAWN FROM EDA

## 5. Payload vs. Orbit Type

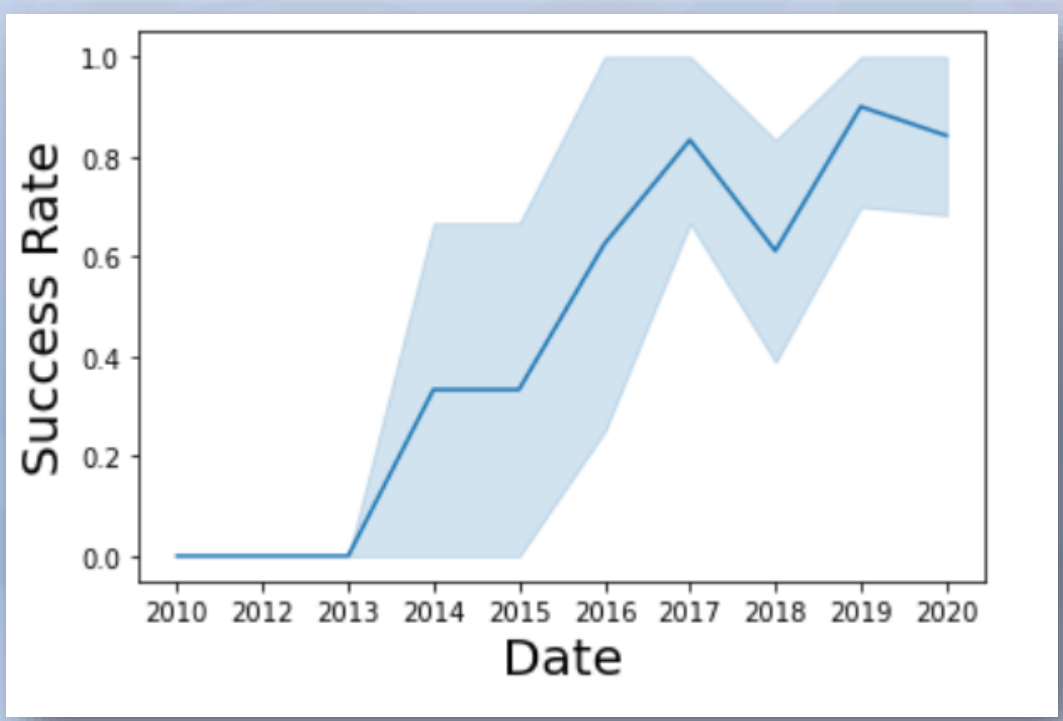
With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS. However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) is apparent.



# INSIGHTS DRAWN FROM EDA

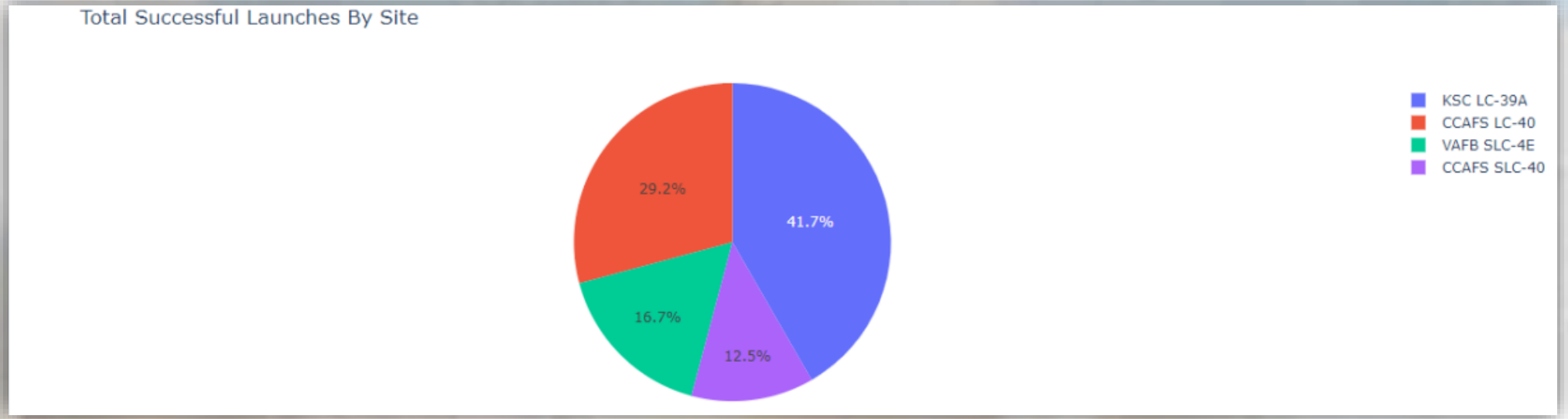
## 6. Launch Success Yearly Trend

It is apparent that the success rate has significantly increased from 2013 to 2020.



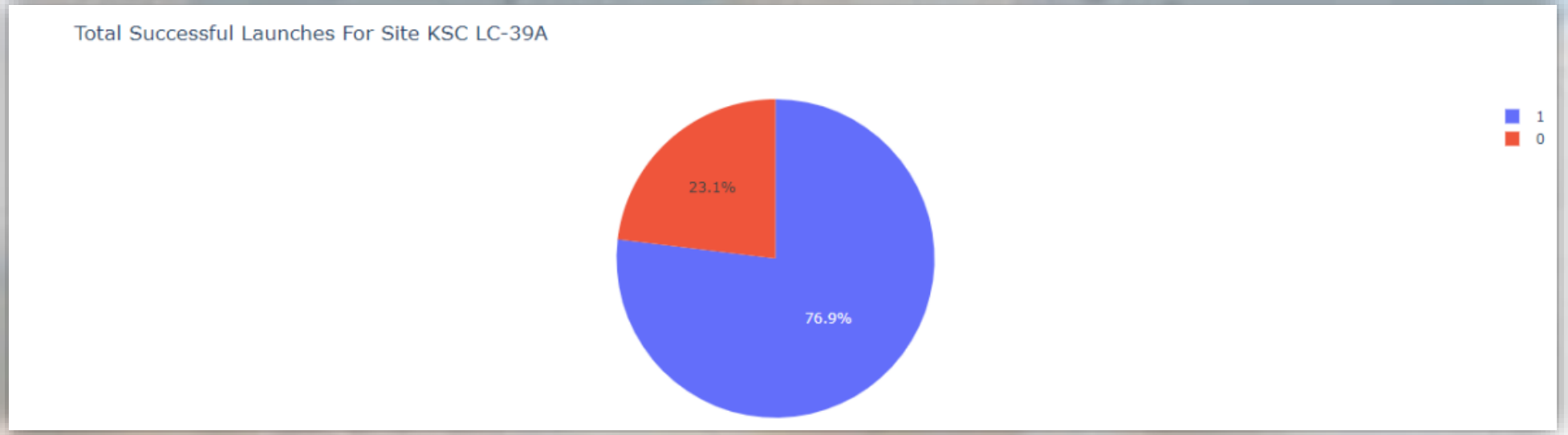
# Building a dashboard with Plotly Dash Successful Launches by Site:

It is observed from the plot that Site KSC LC-39A has the largest successful launches as well the highest launch success rate.



# Building a dashboard with Plotly Dash
# Total Successful Launches for Site KSC LC-39A :

It is observed that 76.9% of the total launches at site KSC LC-39A were successful. This is a the highest success rate of all the different launch sites.




IBM Developer

SKILLS NETWORK

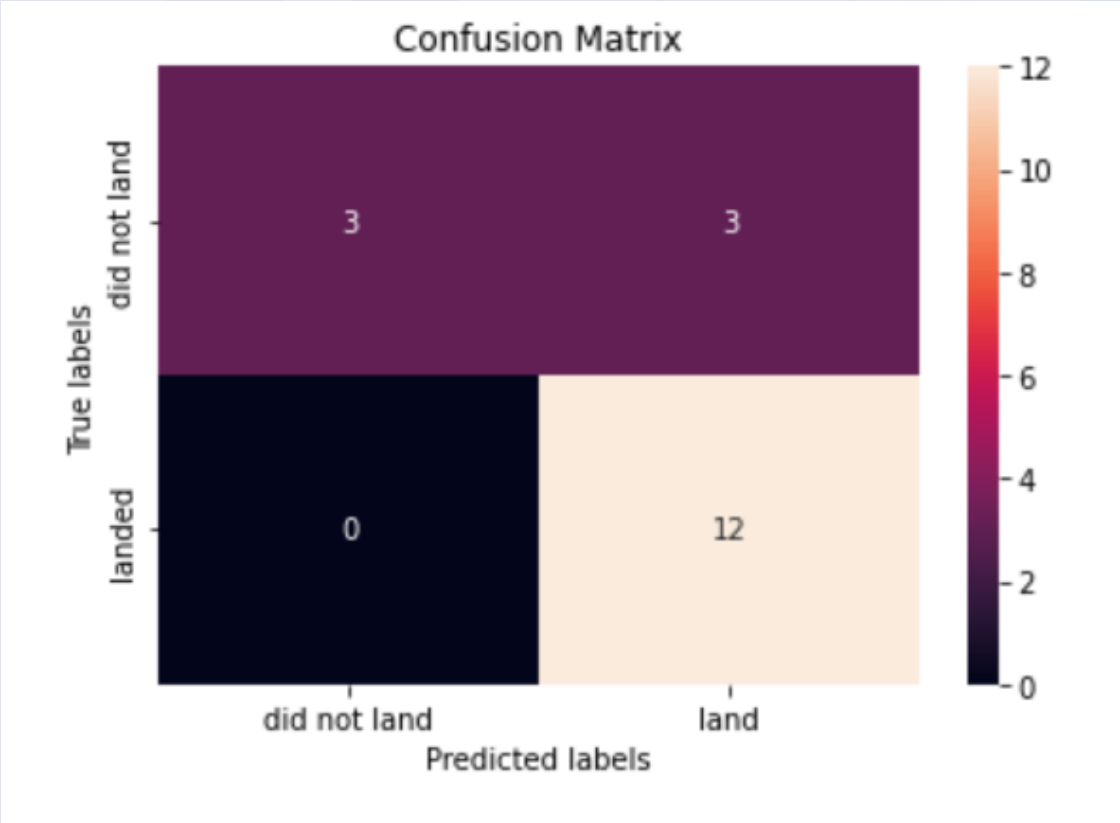
# PREDICTIVE ANALYSIS

It is observed that all the ML models have an accuracy score of 83.33%

Out[33]:

| | ML Method | Accuracy Score (%) |
|---|---|---|
| 0 | Support Vector Machine | 83.333333 |
| 1 | Logistic Regression | 83.333333 |
| 2 | K Nearest Neighbour | 83.333333 |
| 3 | Decision Tree | 83.333333 |

# PREDICTIVE ANALYSIS
## Confusion Matrix:




The chart shows the confusion matrix of the Logistic Regression model that was chosen.
The model only failed to accurately predict 3 labels.

# CONCLUSION

In order to compete with Space X through this process, a general picture of their success methods are:

- All their launch sites are located near the coast, away from nearby cities. This enabled them to test their rocket landings without much interference.
- Site KSC LC-39A had the highest launch success rate out of all the launch sites.
- From 2015 onwards, the success rate of rocket landings significantly increased. It was also apparent that landing success increased with flight number.

All this data was used to train a machine learning model that is able to predict the landing outcome of rocket launches with 83.33% accuracy.


Thank You
IBM Developer
SKILLS NETWORK